文章编号： 1006-3080(2026)02-0276-08 **DOI:** 10.14135/j.cnki.1006-3080.20250803001

水下目标的多尺度上下文感知检测模型

王峻韬[1]， 郑 红[1]， 陆元军[2]， 徐 贤[1]， 吴丽娟[1]

（1. 华东理工大学信息科学与工程学院, 上海 200237； 2. 印孚瑟斯技术(中国)有限公司杭州分公司, 杭州 310056）

摘要: 针对传统模型无法有效处理水下复杂环境噪声、目标尺度变化大、且无法平衡模型大小和精度的问题，本文提出了MSCA-UODA(Multi-scale Context-Aware Underwater Object Detection Algorithm)模型，其设计的上下文增强下采样模块CEADown(Context Enhanced ADown)在降低模型参数量的同时能有效捕获上下文信息，减少了下采样过程中水下环境噪声的影响；同时，提出一种基于双路径部分连接的多尺度特征提取模块CSP-MSPF(Cross Stage Partial-Multi-Scale Partial Feature)，并使用单头注意力机制(Single-Head Self-Attention, SHSA)来改进C2PSA，提高了模型的多尺度特征提取能力。实验表明，相较于基准模型，MSCA-UODA模型在数据集URPC2020和DUO的mAP50分别提升了2.0个百分点和1.1个百分点，参数量下降了12.01%，且综合性能优于目前主流的目标检测模型。

关键词: 水下目标检测；深度学习；注意力机制；下采样；特征提取

中图分类号: TP391.4 **文献标志码:** A

随着海洋经济的快速发展和海洋科学研究的深入推进[1]，对于水下目标检测的精度、效率和鲁棒性提出了越来越高的要求[2]。然而，复杂的水下环境对目标检测技术提出了严峻挑战：光波在水体中的非线性衰减导致图像质量退化，悬浮颗粒物引发的散射效应造成视觉信息失真，加之水下生物群落的动态分布特性，使得传统计算机视觉方法在海洋环境监测中面临显著的技术瓶颈[3]。

目前水下目标检测主要分为基于传统方法的目标检测和基于深度学习的目标检测方法[4]。传统的水下目标检测方法主要基于手工设计的特征，如尺度不变特征变换(SIFT)、加速稳健特征(SURF)等。这些方法在简单的水下环境中可能取得一定的效果，但面对复杂多变的水下场景，其对复杂背景和多尺度目标特征的描述能力有限，检测性能差。随着深度学习技术在计算机视觉领域飞速发展，越来越多的学者应用深度学习来解决目标检测领域的问题[5-6]。当前基于深度学习的目标检测方法主要分为：以R-CNN系列[7]为代表的两阶段检测模型、基于Transformer的DETR系列检测模型[8]以及以SSD[9]、YOLO系列[10]为代表的单阶段检测模型。在水下目标检测领域，杨婷等[11]聚焦于水下图像目标模糊的问题，针对Faster RCNN开展了一系列优化改进，有效提高了水下目标可见度和模型对小目标的检测能力。Song等[12]针对水下图像对比度低、失真等问题提出了全新的二阶段算法Boosting R-CNN，能够在遮挡条件下准确估计对象先验概率。张路等[13]提出使用Fast-EMA模块来替换RT-DERT主干中的BasicBlock，并引入HS-FPN和CGA来改善模型对多尺度目标的漏检和错检问题。Wang等[14]提出了一个DyFishNet和Slim Hybrid编码器来改进DETR，以此更好地提取水下鱼体的纹理特征。Mi等[15]利用改进的白平衡算法和CLAHE提出了一种多尺度融合图像增强算法，并通过Ghost卷积模块、EMA

收稿日期: 2025-08-03

基金项目: 上海市2024年度“科技创新行动计划”(24BC3200500,24BC3200300)

作者简介: 王峻韬(2001—)，男，硕士生，主要研究方向为水下目标检测。E-mail: epic2001@163.com

通信联系人: 郑 红，E-mail: zhenghong@ecust.edu.cn

引用本文: 王峻韬, 郑 红, 陆元军, 等. 水下目标的多尺度上下文感知检测模型 [J]. 华东理工大学学报(自然科学版), 2026, 52(2): 276-283.

Citation: WANG Juntao, ZHENG Hong, LU Yuanjun, *et al*. Multi-Scale Context-Aware Detection Model for Underwater Target[J]. Journal of East China University of Science and Technology, 2026, 52(2): 276-283.

机制和 CARAFE 上采样模块, 改进了 YOLOv5s 水下目标检测算法。谌雨章等[16]结合 SPD-Conv 改进 YOLOv7的头部网络并引入 NAM 注意力机制来检测水下鱼群。

尽管上述研究在水下目标检测领域取得了一定的成果, 但仍存在参数量过大、实时性较低、对水下目标存在漏检、错检等问题。为了解决水下多尺度目标检测精度较低, 特征提取能力较差, 且要保证实时性和轻量化的问题, 本文采用 YOLO11 作为基础模型, 提出了 MSCA-UODA(Multi-Scale Context-Aware Underwater Object Detection Algorithm) 目标检测框架, 能够在减少参数量的同时有效提高检测能力和检测精度: 本文提出了上下文增强下采样结构 CEADown(Context Enhanced ADown), 该模块替换了 YOLO11中原有下采样的卷积模块, 抑制下采样过程中水下环境背景噪声的影响, 降低了模型的参数量和计算量, 提高了其特征检测能力; 设计了一个新的多尺度特征提取模块 CSP-MSPF(Cross Stage Partial-Multi-Scale Partial Feature), 该模块对小部分特征图进行部分多尺度特征提取以提升计算效率, 再利用卷积层结合残差连接融合不同尺度特征, 进而提高模型表达能力; CSPSHSA(Cross Stage Partial with Single-Head Self-Attention) 模块通过引用单头注意力机制 (Single-Head Self-Attention, SHSA) 来改进 C2PSA 模块, 提高了骨干网络的特征提取能力。

1　YOLO11 网络

YOLO11 是 YOLO 算法系列中最新推出的算法之一, 它于 2024 年 10 月被 Ultralytics 提出[17]。作为 YOLO 系列最新的算法, 与之前提出的 YOLO 模型 (如 YOLOv10, YOLOv8) 相比, 它在精度、检测速度和轻量化上有明显提高。

YOLO11 在 YOLOv8 的基础上将模块 C2f 全部替换成 C3K2 模块, 并在模块 SPPF 后提出了 C2PSA 模块, 它在检测头引入了深度可分离卷积模块, 使用该模块替换了检测头中原有的两个常规卷积, 大幅度减少了模型的计算量和参数量, 在保证准确性的同时提高了模型的处理速度。

2　MSCA-UODA 多尺度水下检测算法

2.1　MSCA-UODA 整体结构

本文提出的 MSCA-UODA 的整体架构图如图 1 所示。首先, 模型将 YOLO11 网络中的 Backbone 和 Neck 中的下采样层替换成 CEADown 下采样模块, 该模块基于下采样 ADown[18], 添加了新模块 CEABlock(Context Enhanced Additive Block), 它能在下采样过程中有效减少模型计算量并保留上下文信息, 增强模型的效率和表现, 有效提取图像中关键信息的特征。本文将原有的特征提取模块 C3K2 替换成 CSP-MSPF 模块: 此模块结合 GhostNet 和 FasterNet 的思想, 采用递归式通道划分策略, 选择性地提取并融合不同感受野下的特征表示, 在提高多尺度建模的同时降低了计算量。同时, 通过 1×1 卷积融合不同尺度下的特征、利用残差连接原始输入特征和不同尺度下的特征, 提高了模型的表达能力。最后引入 SHSA注意力机制来改进原有的 C2PSA 模块。

2.2　CEADown 下采样模块

在水下复杂环境中, 由于光学散射和环境背景干扰, 特征在下采样过程中极易出现退化问题。传统下采样方法在降低空间分辨率的同时, 会造成语义与细节信息的部分丢失, 这一问题在水下场景中尤为突出。ADown 通过并行的卷积与池化操作在一定程度上缓解了该问题, 但其仍缺乏全局上下文建模能力, 难以有效抑制复杂背景噪声。

为此, 本文通过融合上下文锚点注意力机制 CAA[19] 和 CAS-ViT[20] 中的 AdditiveBlock 的思想, 设计了一个新的模块 CEABlock。该模块被放在 ADown 无最大池化分支的卷积前面, 旨在控制参数量和计算复杂度的前提下, 减少下采样过程中由噪声引起的特征退化, 并通过上下文增强机制提升模型的特征提取能力, 以更充分地保留水下目标的关键语义信息。

CEABlock 模块的结构包含 3 个协同工作的子模块: 集成器 (Integration)、卷积加性标记混合器 (CATM) 和 CAA, 结构如图 2 所示。Integration由 3 个深度可分离卷积层组成, CATM 模块用于实现高效的自注意力机制, CAA用于捕获上下文信息。该模块相较于 AdditiveBlock在 Integration中使用了 Mish 激活函数来代替 GELU, 并使用 CAA 代替多层感知器(MLP)架构, 其结构如图 2 所示。

相比函数 GELU, Mish 函数具有更平滑的梯度曲线, 使得模型在学习小目标特征时, 能够更好地捕捉细节信息, 而函数 GELU 在某些情况下可能会对小目标的特征进行过度平滑, 导致信息丢失。Mish 函数的连续可微特性使其在反向传播中能保留更完整的梯度信息, 如式(1)所示

$$\mathrm{Mish}(x) = x * \tanh(\ln(1+\mathrm{e}^{x})) \tag{1}$$

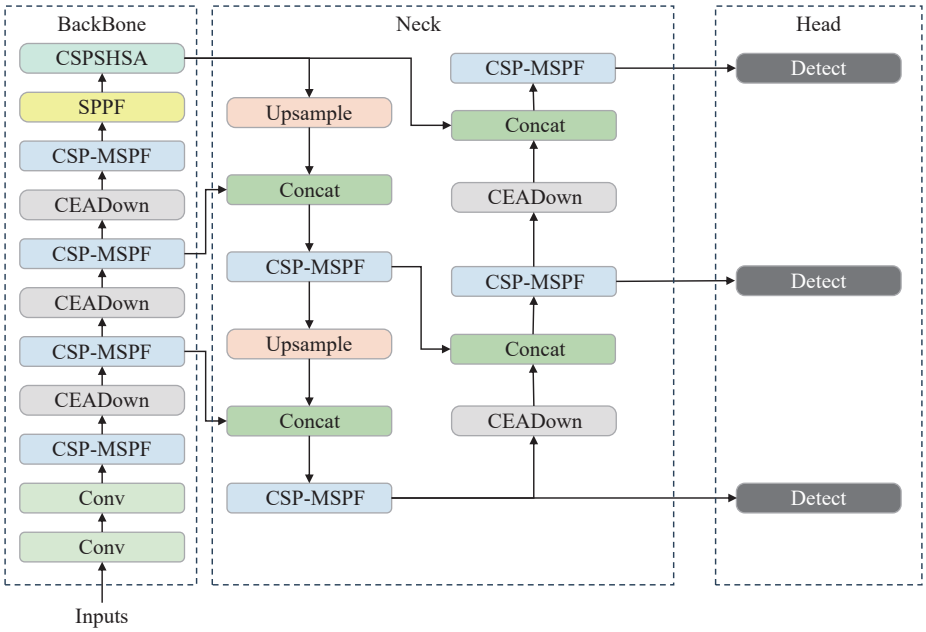


图 1 MSCA-UODA 结构图

Fig. 1 MSCA-UODA structure diagram

通过卷积加性标记混合器 (CATM) 时的特征图输出, 如式 (2) 所示

$$F = L(\psi(\boldsymbol{Q}) + \psi(\boldsymbol{K})) \odot \boldsymbol{V} \tag{2}$$

式中, L 是一个值域在 $R^{N\times d}$ 的函数, 能够对上下文信息进行线性整合; $\psi(x)$ 表示带有 Sigmoid 激活函数的通道注意力和空间注意力的整合; $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别为注意力模型中的查询向量、键向量和值向量, 这三者通过 3 个独立的线性变化得到, 即 $\boldsymbol{Q}=W_q x$, $\boldsymbol{K}=W_k x$,

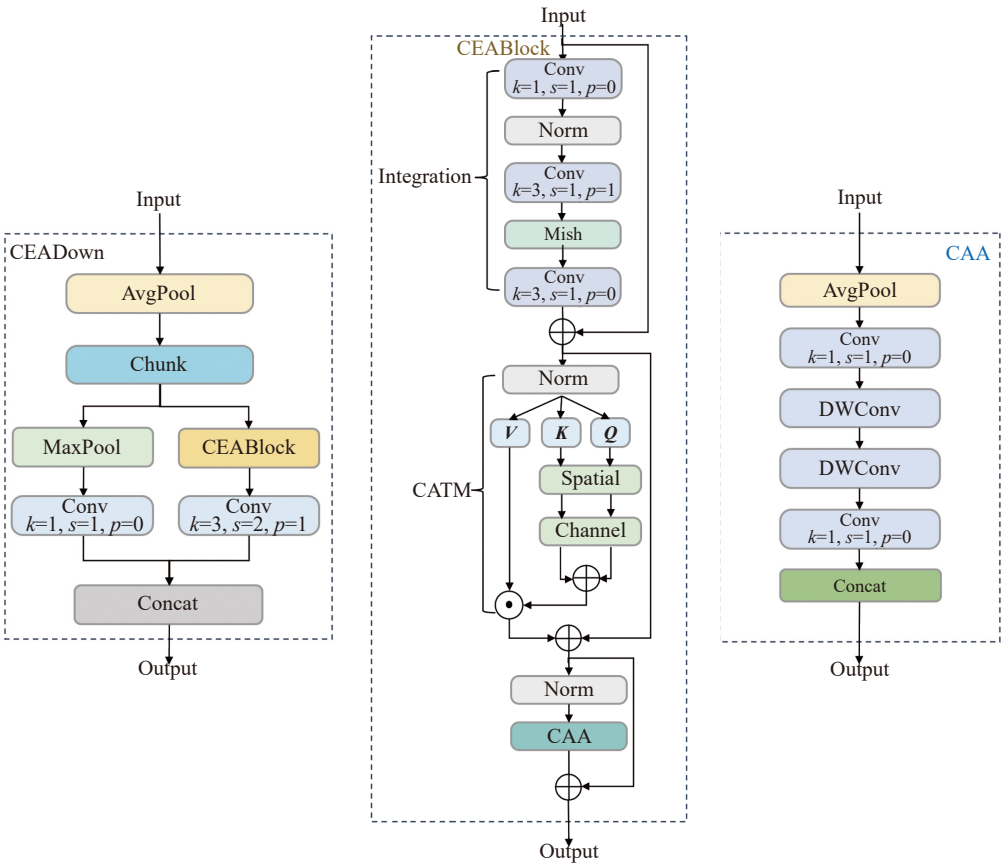


图 2 CEADown 模块框架图

Fig. 2 CEADown module diagram

$V=W_v x$。

AdditiveBlock 原架构中采用了 MLP 结构, 该结构对水下的局部特征进行变换和映射, 但它缺乏有效的全局信息捕获机制。在水下环境中, 噪声通常以复杂且非局部的形式分布, MLP 难以充分利用图像的整体特征信息, 无法有效区分噪声和有用信号, 导致对噪声的鲁棒性不足。

本文采用 CAA 捕获长距离的上下文信息, 其空间建模能力和抗噪声能力均高于 MLP, 同时采用了深度可分离卷积, 使得参数量变少, 它分为 3 个模块: 首先, 对输入特征图 $X\in \mathrm{R}^{C\times H\times W}$ 执行 7×7 平均池化操作和 1×1 的卷积提取, 减少参数量的同时聚合局部信息; 然后, 为建立长距离空间依赖, 使用两个深度可分离卷积在水平和垂直方向上捕捉空间的上下文信息; 最后, 通过一个 1×1 的卷积和 Sigmoid激活函数生成注意力因子并增强。通过这 3 个模块, 能够在特征学习过程中自适应地增强与锚点相关的目标特征, 并抑制无关噪声干扰。复杂背景由于与锚点相关性低而被抑制, 从而显著缓解了水下场景中因光学散射和环境干扰所引起的特征退化问题。

2.3　CSP-MSPF 模块

为了更好地提取多尺度的特征, 本文提出了一个 CSP-MSPF 模块来代替 YOLO11 原本的 C3K2 模块。首先, 基于 FasterNet 提出的 PartialConv 思想[21], 本文中的 MSPF(Multi-Scale Partial Feature)模块对输入通道的一部分进行卷积操作, 减少计算量, 加快了模型推理速度; 其次, 基于 GhostNet 的思想[22], 先对特征图进行一次卷积, 然后根据得到的特征图对每个通道进行特征映射, 最终生成的特征图与采用传统卷积得到的特征图具有相同的通道数, MSPF 模块的主要结构如图 3 所示。

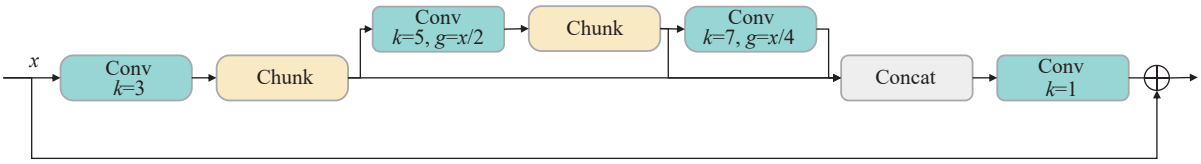


图 3　MSPF 模块结构图

Fig. 3　MSPF module structure diagram

CSP-MSPF 的 MSPF 模块主要包含 4 个卷积操作: 首先, 输入的特征图 $x\in \mathrm{R}^{\mathrm{B}\times C\times H\times W}$ 经过 3×3 的卷积, 并沿其输出维度均分为 $Y1$ 和 $Y2$, 如式 (3) 所示; 然后, 将特征图 $Y1$ 通过一个卷积核为 5 的分组卷积, 再将特征图沿通道维度平均分成为 $Y3$ 和 $Y4$, 如式 (4) 所示; 再次, 将特征图 $Y3$ 通过一个卷积核为 7 的分组卷积得到特征图 $Y5$, 如式 (5) 所示; 最后, 将特征图 $Y2$、$Y4$ 和 $Y5$ 沿通道维度拼接, 通过一个 1×1 卷积和残差连接得到最终的特征图, 如式 (6) 所示。通过以上操作, MSPF 模块利用分通道并行与层级拆分的方法, 从多尺度上提取并融合特征; 并行分支使用不同的分组和卷积核尺寸, 捕获从局部到全局的多样化特征; 逐级拆分与跳跃连接增强信息复用与梯度流动。最终, 1×1 卷积实现跨通道融合, 并通过跳跃连接与原始输入相加, 既保持特征完整性, 又提升模块与网络的协同性, 同时显著降低计算复杂度。

$$Y1, Y2 = \mathrm{Split}(\mathrm{Conv}_{3\times 3}\left(x^{B\times C\times H\times W}\right), \mathrm{dim}=1, \mathrm{chunks}=2) \tag{3}$$

$$Y3,\ Y4 = \mathrm{Split}(\mathrm{Conv}_{5\times 5}\left(Y1,\ \mathrm{group}=\frac{C}{2}\right),\ \mathrm{dim}=1,\ \mathrm{chunks}=2) \tag{4}$$

$$Y5 = \mathrm{Conv}_{7\times 7}\left(Y3,\ \mathrm{group}=\frac{C}{4}\right) \tag{5}$$

$$\mathrm{OutPut} = \mathrm{Conv}_{1\times 1}(\mathrm{Concat}(Y5,\ Y4,\ Y2,\ \mathrm{dim}=1)) + x^{B\times C\times H\times W} \tag{6}$$

CSP-MSPF 模块通过集成多个 MSPF 模块, 替换 C2f 中原有的 Bottleneck 模块, 运用了多个不同大小的卷积核, 可以全面理解图像信息, 不同尺度的特征包含了不同层次的信息, 将这些特征融合可使模型对光照变化、物体遮挡等复杂场景下物体的变化有更好的适应性。通过使用分组卷积, 减少了模型的参数量和计算量, 在保证模型性能的前提下, 减少计算资源的消耗, 提高了模型实时检测能力。

2.4　CSPSHSA 模块

YOLO11 在 Backbone 中提出了一个新的模块 C2PSA, 该模块放在 SPPF 模块后, 主要用于增强主干网络的特征提取功能。C2PSA 通过在通道和空间维度上进行注意力机制融合, 提升了特征图的表达能力: 首先对特征图的通道进行自适应加权, 然后结合空间注意力进一步调整重要区域的响应, 从而增强目标检测的精度。然而, C2PSA 仍存在一定的局限性: 自注意力分支在高分辨率特征图下会引入额外的参数量与计算开销; PSA 注意力的复杂性增加

了模型的参数量,进而提高了训练和推理过程中的资源消耗。为了更好降低模型的计算和内存开销,提高模型检测精度,本文提出了CSPSHSA模块。

单头注意力机制(Single-Head Self-Attention, SHSA)[23]是一种高效的注意力机制,其原理是通过对部分通道进行单头注意力机制的应用,减少一部分计算的冗余和对内存访问的成本,最终将投影应用于全部的通道,保障注意力的有效性,其结构图如图4所示,其中C为通道数,p为通道分割比例。在CSPSHSA中,SHSA被嵌入CSP框架中,用于对部分特征分支进行全局建模。具体而言,CSP结构首先在通道维度上将输入特征划分为两支,其中一支引入SHSA对语义密集的子通道执行单头自注意力建模,从而有效捕获长程空间依赖;另一支则保留原始特征传递。随后,两支特征在通道维度上进行融合,实现全局上下文信息与局部表征的互补集成。CSPSHSA在保持CSP梯度稳定与参数效率优势的同时,引入SHSA注意力,降低计算冗余,以较低复杂度增强特征表达,满足实时目标检测的效率与性能需求。

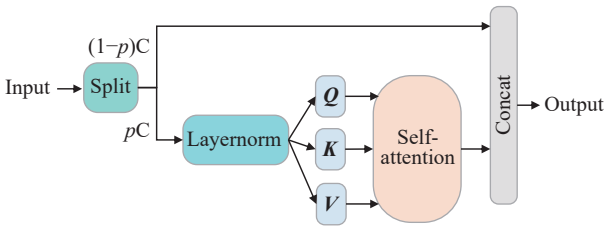


图4 SHSA结构图

Fig. 4 SHSA structure diagram

3 实验结果及分析

3.1 数据集及运行环境

本文选取的数据集为URPC2020和DUO数据集[24]。URPC2020数据集是拍摄于真实的海底环境,包含了海星、扇贝、水草和海参这4类水下生物的图片,总共5543张图片,数据集按8∶1∶1的比例随机划分成训练集、验证集与测试集。DUO数据集为URPC挑战赛多年数据集的整合,其筛除了重复数据,并对部分错误的数据集进行重新标注,总共7782张图片。为了更好地评估模型泛化能力,本文按照7∶1∶2的比例随机将DUO数据集划分为训练集、验证集和测试集。

本实验使用的硬件包括Intel(R) Xeon(R) Platinum 8255C的CPU, RTX3090(24 G)的GPU, 43 G内存,Linux操作系统等来实现,软件运行环境为PyTorch2.0.0, Python3.8, CUDA 11.8,输入图像分辨率为640×640。对于URPC2020数据集,批量大小设置为16,迭代次数设置为150次。对于DUO数据集,批量大小设置为32,迭代次数设置为200次。

3.2 评价指标

为了系统地评估本算法前后的性能差异,研究采用准确率(P)、召回率(R)、平均精度均值(mAP)和参数量(Parameter)作为核心指标来客观评价模型性能,表达式如式(7)~(9)所示。

$$P=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{7}$$

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{8}$$

$$\mathrm{mAP}=\frac{\sum_{j=1}^{k}\mathrm{AP}(j)}{k} \tag{9}$$

式中,TP指模型正确预测的正样本数量,FP为模型误判为目标的错误检测数量,FN是实际存在但未被模型检出的目标数量。对于第j类目标,其平均精度为AP(j),数据集的类别数量由k表示。

3.3 实验结果

3.3.1 消融实验 本文采用多种方式来改进YOLO11算法,为了验证各个模块的有效性,本文在URPC数据集上进行了消融实验,结果如表1所示。

从表1可以得知,相较于基准模型YOLO11,改进的各个模块在召回率和mAP50均有提高,且各个模块之间无相互冲突。基准模型通过添加CEADown下采样模块后,参数量下降12.4%,准确率上升0.2个百分点,召回率提高0.9个百分点,mAP50提高1.1个百分点,说明CEADown下采样模块能够在有效减少模型参数量的同时大幅增强了模型的特征提取能力。通过引入CEADown下采样模块和CSP-MSPF模块,模型的召回率提高到78%,mAP50进一步提高到84.8%,表明模型在多尺度特征的信息选择能力增强,能更好地捕获多尺度信息。将CEADown, CSP-MSPF和CSPSHSA相结合,得到了本文所提出的MSCA-UODA模型,其参数量减少12.01%,准确率提高了0.9个百分点,召回率提升1.6个百分点,mAP50提升2个百分点,在降低模型复杂度和存储空间,减少计算资源的消耗的同时,提高了模型的检测能力。

表2所示为CEADown下采样模块的消融实验,原始的ADown模块虽然有效降低了模型的参数量和计算量,但准确率、召回率和mAP50都有不同程度的下降。通过结合本文提出的CEABlock模块,下

表 1　模型各个模块的消融实验

Table 1　Ablation experiments for each module of the model

Number	YOLO11	CEADown	CSP-MSPF	CSPSHSA	Parameter	*P*/%	*R*/%	mAP50/%
1	√				2.58×10^6	82.8	75.8	83.1
2	√	√			**2.26×10^6**	83.0	76.7	84.2
3	√		√		2.63×10^6	83.2	76.6	83.8
4	√			√	2.55×10^6	83.0	77.2	83.3
5	√	√	√		2.30×10^6	82.4	**78.0**	84.8
6	√	√	√	√	2.27×10^6	**83.7**	77.4	**85.1**

表 2　CEADown 消融实验

Table 2　CEADown ablation experiment

Method	Parameter	*P*/%	*R*/%	mAP50/%	FLOPs/GFLOPs
YOLO11	2.58×10^6	82.8	75.8	83.1	6.3
YOLO11+ADown	**2.10×10^6**	82.5	75.9	82.9	**5.3**
YOLO11+ADown+CEABlock(CEADown)	2.26×10^6	**83.0**	**76.7**	**84.2**	6.1

采样 CEADown 模块的计算量和参数量总体低于 YOLO11 模型本身, 且准确率、召回率和 mAP50 都明显超过了 YOLO11, 证明了 CEABlock 模块的有效性。

为了更加直观体现模型和模块的能力, 图 5 展示了不同情况下的热力图。相较于基线模型, 添加了 CEADown 模块的模型在图片模糊、低对比度和多目标背景这几种不同水下的情况下, 能够有效减少背景噪声的影响。同时, 本文提出的 MSCA-UODA 模型能在 CEADown 的基础上提取关键目标信息, 增强特征提取的准确率。消融实验验证了本文算法各模块的有效性, 热力图分析直观展现了其特征捕捉优势, 综合表明本文算法在各项指标上均优于基准模型。

3.3.2　不同模型对比实验　为评估所提方法的有效性, 在统一的硬件配置和测试数据集条件下开展对比测试, 结果如表 3 所示。由表 3 可以得出, 本文改进后的算法相较于目前先进的目标检测算法均有提升, 且参数量级相对较小, 相较于基准模型 YOLO11n, 本文模型参数量降低 12.01%; URPC2020 和 DUO 数据集的 mAP50 分别提升了 2.0 个百分点和 1.1 个百分点。本文提出的模型相较于二阶段 Faster-RCNN 参数量减少 94.51%, 在 URPC2020数据集上准确率提高 2.9 个百分点, 召回率提高 1 个百分点, mAP50 提高 1.1 个百分点, mAP50-95 提高 2.4 个百分点; 在 DUO 数据集, 准确率提高 1.3 个百分点, 召回率提高 0.6 个百分点, mAP50 提高 1.3个百分点, mAP50-95 提高 0.6 个百分点。与 YOLOv12n模型[25] 相比, 本文模型在 URPC2020 数据集参数量减少 10.98%, 准确率提高 0.9 个百分点, 召回率提高 2.1 个百分点, mAP50提高 2.1个百分点, mAP50-95 提高 1.5 个百分点, 同时在 DUO 数据集上本文提出的模型在召回率、mAP50和 mAP50-95 有较大提高。这说明本文算法在平衡模型大小和模型精度上的改进与对多尺度目标检测的能力提升方面是有效的。除了上述模型外, 本文还将所提方法与当前广泛应用的多个 YOLO 系列 Nano 版本的模型进行对比, 具体包括 YOLOv5n、YOLOv7-tiny、YOLOv8n及YOLOv10n。结

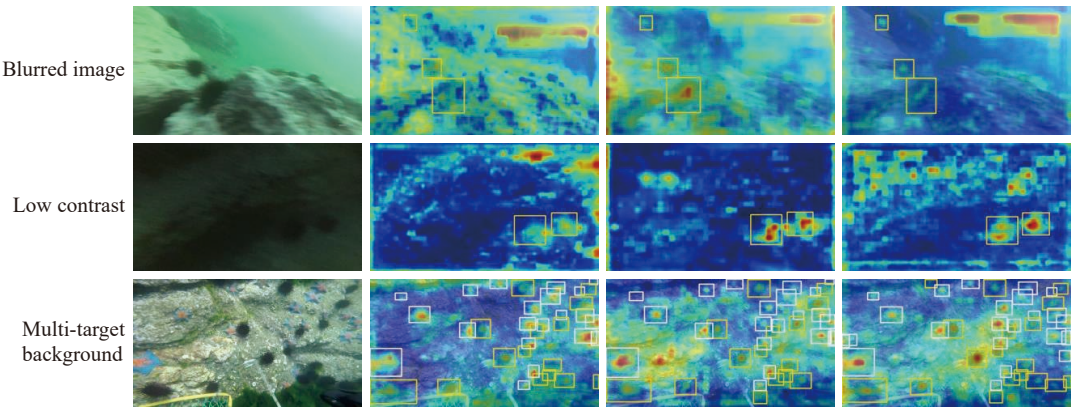


图 5　不同水下环境的热力图

Fig. 5　Heat maps of different underwater environments

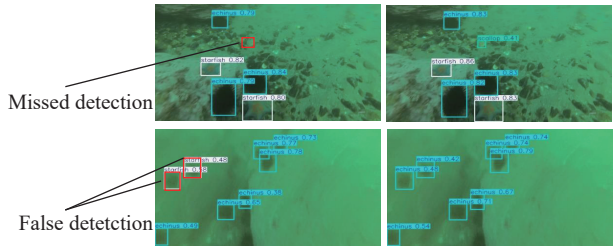
表 3 不同模型的对比实验

Table 3 Comparative experiments of different models

Model	Parameter	URPC2020				DUO			
		P/%	*R*/%	mAP50/%	mAP50-95/%	*P*/%	*R*/%	mAP50/%	mAP50-95/%
Faster R-CNN	41.36×10^6	80.8	76.4	84.0	48.7	84.8	74.4	83.4	64.8
YOLOv5n	1.76×10^6	84.7	75.8	82.6	46.9	83.9	74.8	82.2	58.4
YOLOv7-tiny	6.02×10^6	82.3	76.6	82.9	45.3	86.2	75.5	84.6	62.3
YOLOv8n	3.00×10^6	80.5	76.3	82.6	49.4	82.4	75.6	83.0	63.0
YOLOv10n	2.65×10^6	82.3	76.1	83.1	49.2	84.3	73.6	82.4	62.9
YOLO11n	2.58×10^6	82.8	75.8	83.1	49.6	84.5	74.9	83.6	64.2
YOLOv12n	2.55×10^6	82.8	75.3	83.0	49.6	86.3	73.1	83.4	63.6
Ours	2.27×10^6	83.7	77.4	85.1	51.1	86.1	75.0	84.7	65.4

合表 3 所示结果可知，本文提出的模型在 mAP50 和 mAP50-95 精度最高，同时参数量仅略高于 YOLOv5n。这表明本文方法在显著提高水下目标检测精度的同时，有效控制了模型复杂度，实现了精度与轻量化程度的较优平衡。

图 6 示出了水下检测结果的对比图。通过对比可以得出，由于水下复杂环境，目标对象较小等因素，YOLO11 会出现漏检、错检的情况，而本文提出的模型可以有效解决该问题。



(a) YOLO11 (b) MSCA-UODA

图 6 水下检测结果

Fig. 6 Underwater detection results

4 结束语

针对水下环境复杂、目标尺度变化大，传统算法无法平衡检测精度和模型大小的问题，本文基于 YOLO11 算法提出了 MSCA-UODA 模型。首先，通过使用 CEADown 模块减少了模型的参数量，降低模型下采样过程中受到的背景噪声干扰，提高了模型对水下特征的上下文提取能力。其次，通过使用 CSP-MSPF 模块来提高模型的鲁棒性和多尺度特征的检测能力。最后，使用 CSPSHSA 模块减少计算冗余，提高模型检测精度。本文提出的算法在多个实验上证明了其有效性，但其仍可以从计算量和检测速度等指标上进行深入研究，考虑结合图像增强算法等处理方法，提升其在水下多尺度复杂环境的检测能力。

参考文献：

[1] 刘飞, 杨德刚, 章鑫, 等. 基于 YOLOv8 改进的水下目标检测算法 [J]. 计算机与现代化, 2025(1): 113-119.

[2] CHEN X, YUAN M, YANG Q, *et al*. Underwater-YCC: Underwater target detection optimization algorithm based on YOLOv7[J]. Journal of Marine Science and Engineering, 2023, 11(5): 995.

[3] 张明华, 刘佳艺, 石少华, 等. 基于全局特征提取和提示学习的水下图像增强 [J]. 华中科技大学学报 (自然科学版), 2025, 53(3): 31-40.

[4] 李康. 基于可变形卷积神经网络的海洋柔性生物目标检测算法研究 [D]. 哈尔滨: 哈尔滨工程大学, 2023.

[5] 齐鑫伟, 侍洪波, 宋冰, 等. 基于自上而下注意力机制的零样本目标检测 [J]. 华东理工大学学报 (自然科学版), 2024, 50(6): 859-868.

[6] 李耀, 李梅. 基于多尺度 WideResNet 的铁轨缺陷小样本检测算法 [J]. 华东理工大学学报 (自然科学版), 2025, 51(6): 843-849.

[7] GIRSHICK R, DONAHUE J, DARREL T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, Ohio, USA: IEEE Computer Society, 2014: 580-587.

[8] CARION N, MASSA F, SYNNAEVE G, *et al*. End-to-end object detection with transformers[C]//European Conference on Computer Vision. CHAM: Springer International Publishing, 2020: 213-229.

[9] LIU W, ANGUELOV D, ERHAN D, *et al*. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference. Amsterdam, Netherlands: Springer International Publishing, 2016: 21-37.

[10] REDMON J, DIVVALA S, GIRSHICK R, *et al*. You only look once: Unified, real-time object detection[C]//Proceed-

ings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 779-788.

[11] 杨婷, 高武奇, 王鹏, 等. 自动色阶与双向特征融合的水下目标检测算法 [J]. 激光与光电子学进展, 2023, 60(6): 132-143.

[12] SONG P, LI P, DAI L, *et al*. Boosting R-CNN: Reweighting R-CNN samples by RPN's error for underwater object detection[J]. Neurocomputing, 2023, 530: 150-164.

[13] 张路, 魏本昌, 魏鸿奥, 等. 基于改进 RT-DETR 的水下目标检测 [J]. 计算机系统应用, 2024, 33(12): 131-140.

[14] WANG Z, RUAN Z, CHEN C. DyFish-DETR: Underwater fish image recognition based on detection transformer[J]. Journal of Marine Science and Engineering, 2024, 12(6): 864: 1-15.

[15] MI Y , CHI M , ZHANG Q , *et al*. Research on multi-scale fusion image enhancement and improved YOLOv5s lightweight ROV underwater target detection method[J]. Scientific Reports, 2024, 14(1): 39550442.

[16] 谌雨章, 王诗琦, 周雯, 等. 基于 SPD-Conv 结构和 NAM 注意力机制的鱼群小目标检测 [J]. 计算机科学, 2024, 51(S1): 438-444.

[17] KHANAM R, HUSSAIN M. Yolov11: An overview of the key architectural enhancements[J/OL]. (2024-10-23). https://www.arxiv.org/abs/2410.17725.

[18] WANG C Y, YEH I H, MARK LIAO H Y. Yolov9: Learning what you want to learn using programmable gradient information[C]//European Conference on Computer Vision. Switzerland: Cham Springer, 2024: 1-21.

[19] CAI X, LAI Q, WANG Y, *et al*. Poly kernel inception network for remote sensing detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2024: 27706-27716.

[20] ZHANG T, LI L, ZHOU Y, *et al*. Cas-vit: Convolutional additive self-attention vision transformers for efficient mobile applications[J/OL]. (2024-08-07). https://arxiv.org/abs2408.03703, 2024.

[21] CHEN J, KAO S, HE H, *et al*. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE, 2023: 12021-12031.

[22] HAN K, WANG Y, TIAN Q, *et al*. Ghostnet: More features from cheap operations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [s.l]: IEEE, 2020: 1580-1589.

[23] YUN S, RO Y. Shvit: Single-head vision transformer with memory efficient macro design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [s.l.]: IEEE Computer Society, 2024: 5756-5767.

[24] 罗逸豪, 刘奇佩, 张吟, 等. 基于深度学习的水下图像目标检测综述 [J]. 电子与信息学报, 2023, 45(10): 3468-3482.

[25] TIAN Y, YE Q, DOERMANN D. Yolov12: Attention-centric real-time object detectors[J/OL]. (2025-02-18). https://arXiv.org/abs/2502.12524.

Multi-Scale Context-Aware Detection Model for Underwater Target

WANG Juntao[1], ZHENG Hong[1], LU Yuanjun[2], XU Xian[1], WU Lijuan[1]

(1. School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China; 2. Infosys Technologies (China) Co. Ltd, Hangzhou Branch, Hangzhou 310056, China)

Abstract: To address the limitations of traditional models in handling complex underwater environmental noise, large variations in target scale, and the trade-off between model size and accuracy, the MSCA-UODA (Multi-scale Context-Aware Underwater Object Detection Algorithm) was proposed. The model includes a context-enhanced downsampling module, CEADown (Context-Enhanced ADown), which effectively reduces model parameters, captures contextual information efficiently, and mitigates underwater environmental noise. Additionally, it introduces a multi-scale feature extraction module based on dual-path partial connection, named CSP-MSPF (Cross Stage Partial-Multi-scale Partial Feature), and incorporates the SHSA (Single-Head Self-Attention) mechanism to enhance the C2PSA module, thereby improving the model's multi-scale feature extraction capability. Experimental results show that on the URPC2020 and DUO datasets, MSCA-UODA improved mAP50 by 2.0 percentage points and 1.1 percentage points, respectively, compared to the baseline model, while reducing the number of parameters by 12.01%. Its overall performance surpassed that of current mainstream object detection models.

Key words: underwater object detection; deep learning; attention mechanism; downsampling; feature extraction

（责任编辑：王晓丽）